RESTRICTED-RECOURSE BOUNDS FOR STOCHASTIC LINEAR PROGRAMMING

DAVID P. MORTON

Graduate Program in Operations Research, The University of Texas at Austin, Austin, Texas 78712, morton@mail.utexas.edu

R. KEVIN WOOD

Operations Research Department, Naval Postgraduate School, Monterey, California 93943, kwood@nps.navy.mil

We consider the problem of bounding the expected value of a linear program (LP) containing random coefficients, with applications to solving two-stage stochastic programs. An upper bound for minimizations is derived from a restriction of an equivalent, penalty-based formulation of the primal stochastic LP, and a lower bound is obtained from a restriction of a reformulation of the dual. Our "restricted-recourse bounds" are more general and more easily computed than most other bounds because random coefficients may appear anywhere in the LP, neither independence nor boundedness of the coefficients is needed, and the bound is computed by solving a single LP or nonlinear program. Analytical examples demonstrate that the new bounds can be stronger than complementary Jensen bounds. (An upper bound is "complementary" to a lower bound, and vice versa). In computational work, we apply the bounds to a two-stage stochastic program for semiconductor manufacturing with uncertain demand and production rates.

This paper develops new techniques for bounding the expected value of a stochastic linear program, which is a linear program (LP), some or all of whose coefficients are random. The random coefficients may be discretely or continuously distributed, may be independent or contain dependencies, and may occur anywhere in the objective function, right-hand side, or constraint matrix. Calculating (or estimating) the expected value of a stochastic LP is key to solving two-stage and multi-stage stochastic programs with recourse (Dantzig 1955).

It is usually impossible to compute exactly the expected value of a stochastic LP unless the random coefficients are discretely distributed and the total number of realizations of the coefficients is small, or the problem has a very special structure. As a result, solution techniques for stochastic programming with recourse usually involve approximations that are based on Monte Carlo sampling (e.g., Ermoliev 1983, Dantzig and Glynn 1990, Higle and Sen 1991) or on deterministically valid bounds. This paper involves approximations of the latter type.

When initial lower and upper bounds are insufficiently tight, they can often be improved within a sequential-approximation algorithm that iteratively partitions the support of the random variables (e.g., Kall et al. 1988). We do not discuss the details of these algorithms here, but note that our restricted-recourse bounds can be incorporated within such algorithms just as well as other bounds can, and possibly better because of improved computational efficiency and fewer technical requirements.

The problem of computing the expected value of a stochastic LP also arises as a stand-alone problem. For instance, the stochastic maximum-flow problem (e.g., Evans 1976) calculates the expected value of the maximum flow through a network whose arc capacities are nonnegative random variables. Another example is the stochastic PERT problem (Fulkerson 1962), which evaluates the expected length of a longest path in a directed acyclic network with stochastic arc lengths. Both of these problems can also appear as second-stage *recourse problems* in two-stage stochastic programs with recourse. In the former case, first-stage decisions make resource-constrained investments in arc capacities; the problem is to maximize the expected value of the second-stage *recourse function*, i.e., the expected value of the maximum flow (Wallace 1987a, Wollmer 1991). In the latter case, recourse-constrained investments reduce the (expected) lengths of certain arcs for the purpose of minimizing the expected length of a maximum-length path (Wollmer 1985).

Stochastic-programming formulations are used to model many real-world planning problems that are similar to, but more complex than the simple network problems just described: The structure of the constraints is usually more involved, and random coefficients may occur anywhere in the objective, right-hand, side or constraint matrix. Applications include capacity-expansion planning in electric power systems (Dantzig et al. 1989) and the telecommunications industry (Sen et al. 1994), finance (Cariño et al. 1994), hydroelectric scheduling (Jacobs et al. 1995), forest harvest management (Gassmann 1989), and vehicle allocation (Powell 1986). We regard the bounds developed in this paper as tools with potential for significantly increasing the applicability of sequential-approximation methods for solving both the simple and complex problems mentioned above.

Deterministic bounds for stochastic LPs may be derived either from functional approximations of the recourse function or from approximations of the probability distribution that governs the random parameters. The most widely known bound is based on Jensen's inequality for convex functions and, interestingly, this bound can be derived from either type of approximation (e.g., Kall and Wallace 1994, §3.4). The bound is easy to compute—the recourse function is evaluated once at the mean of the random parameters—but the requirement for convexity (or concavity) limits its applicability.

Piecewise-linear bounds (Birge and Wallace 1988, Wallace 1987b, Birge and Wets 1986, Birge and Wets 1989) are based on a functional approximation. These bounds can be computed by solving a sequence of restrictions of the recourse problem and thus yield upper bounds for a minimization problem (Birge and Wallace 1988). For the purpose of solving stochastic LPs, it is significant that the piecewise-linear bound is complementary to the Jensen bound. That is, when Jensen's inequality may be applied to yield a lower bound, the piecewise-linear bound is an upper bound and vice versa. The bounds are derived from a separable approximation of the recourse function that contains a sum of r one-dimensional expectations, where r is the number of stochastic parameters. Bounds with this structure typically can be applied to problems with random parameters that have bounded or unbounded support. The bounds are relatively easy to compute—they require $\mathcal{O}(r)$ solutions of the recourse problem—but they are effectively limited to problems with stochastic right-hand sides only. Other examples of bounds based on functional approximations can be found in Birge (1985), Edirisinghe and Ziemba (1992), and Powell and Frantzeskakis (1994). We discuss these in more detail after looking at distributional approximations.

Bounds derived from distributional approximations are exemplified by the Edmundson–Madansky (E–M) bound (Edmundson 1956, Madansky 1959) and its generalizations. The E–M bound is constructed by replacing the original probability distribution with an approximating distribution that has mass only at the extreme points of the random parameters' support. This bound can be stronger than alternatives such as the piecewise-linear bound (Wallace 1987b, Birge and Wallace 1988), but it has significant drawbacks: (a) The random parameters must be independent and bounded, (b) the recourse function must be convex or concave, and (c) the recourse function must be evaluated $\mathcal{O}(2^r)$ times. Gassmann and Ziemba (1986) generalize the E–M bound to dependent random variables defined on polyhedral sets that need not be bounded. Frauendorfer (1988) generalizes the E–M bound for dependent random variables defined on (multidimensional) rectangles by incorporating cross-moment information.

Frauendorfer (1992) develops a sequential-approximation algorithm that may avoid the exponential effort normally required for bound computation by applying the bounds on a Cartesian product of simplices. The approach is applicable to LPs with possibly dependent stochastic right-hand sides

and objectives. Edirisinghe and Ziemba (1994a, 1994b) extend the bounds to the convex–concave case by applying them over more general, possibly unbounded, polyhedra. Edirisinghe (1996) and Edirisinghe and You (1996) develop distributional bounds on simplices and show how to tighten these bounds by using second-order information and by using simplices with least possible volume. See Dulá (1992) for related work using a functional approximation.

Conditional probability mass and moment calculations—these are required in a sequential-approximation algorithm—can be difficult when distributional bounds are applied on simplices or more general polyhedra that are "unnatural" with respect to the underlying probability model. It is possible to circumvent these difficulties (Edirisinghe 1996, Edirisinghe and You 1996, Frauendorfer 1992) by generating manageable numbers of scenarios via Monte Carlo sampling and by applying the bounds to the resulting empirical model rather than the original model. For large sample sizes, this approach is justified by the theory of epi-convergence (King and Wets 1991).

We note that several of the extensions of the Jensen and E–M bounds, as well as some bounds based on functional approximations (e.g., Birge and Dulá 1991, Birge and Wets 1986) may be derived in an alternative manner—in particular, as solutions of generalized moment problems (Birge and Wets 1987; Kall 1988; Frauendorfer 1992; Edirisinghe and Ziemba 1994a, 1994b). This interpretation of bounds for stochastic programs begins with the work of Dupačová in a game-theoretic setting (Žáčková 1966) and also has applications in models with incomplete information regarding the probability distribution (Dupačová 1994).

The restricted-recourse bounds proposed in this paper are attractive because they: (a) are relatively easy to compute, (b) are deterministically valid, (c) have no restrictions as to where random coefficients may occur in the problem, and (d) are computable for random coefficients with bounded or unbounded ranges. In order to apply the restricted-recourse methodology, we require that bounds on optimal dual variables (for the primal bound) or primal variables (for the dual bound) of the recourse problem be specified. In practice, specifying bounds on primal variables is usually simple, and models that contain explicit penalties for constraint violations automatically incorporate bounds on dual variables. We will illustrate that special structures may be exploited to derive appropriate dual bounds in many of the cases, too.

The restricted-recourse bounds in this paper may be viewed as improvements on the penalty-based "aggregation bounds" developed by Birge (1985) for stochastic LPs and by Edirisinghe and Ziemba (1992) for two-stage stochastic (convex) programs. The aggregation bounds are based on functional approximations for recourse problems that have explicit penalties for constraint violations and that have random coefficients in the right-hand side only. With this structure, the expected value of the penalty term separates into a sum of one-dimensional expectations that is computed for fixed first- and second-stage decision variables. Because the bounds are based on restrictions, they

are upper (lower) bounds for minimization (maximization) problems.

Powell and Frantzeskakis (1994) first use the term "restricted recourse" for bounding two-stage (and multi-stage) network recourse problems; their scope is limited to network models with stochastic right-hand sides and constraint-violation penalties. Like aggregation bounds, their methods lead to separable approximations of the recourse function. An important difference is that for any first-stage decision, Powell and Frantzeskakis compute the expected cost of a single best second-stage decision, and they minimize this expected cost over all first-stage decisions. As a result, this model is less restricted than that employed by aggregation bounds, and tighter bounds can be obtained. (We note also that Vladimirou and Zenios (1997) use "restricted recourse" to refer to a class of models designed to control the variability of recourse decisions in two-stage stochastic linear programs.)

Our restricted-recourse bounds substantially generalize the bounds of Powell and Frantzeskakis because (a) our bounds are applicable to general LPs, (b) random coefficients may occur anywhere in the problem, and (c) we describe both primal and dual bounds.

Our first application of the generalized version of restricted-recourse was to a stochastic network-interdiction problem (Cormican et al. 1996). We solve that problem using a sequential-approximation algorithm that incorporates a Jensen upper bound and a restricted-recourse lower bound. Restricted recourse was not explicitly used to derive the bound, however, because a reformulation and an application of Jensen's inequality to that reformulation suffices. Nonetheless, the current paper generalizes the basic technique: If bounds on optimal dual variables are available, we show that any stochastic LP can be converted to an equivalent problem with a stochastic objective function but deterministic constraints. By solving a restriction of the reformulation in which a single recourse decision is allowed, an upper (lower) bound results when the original problem is a minimization (maximization). We classify this bound as a *primal restricted-recourse* bound. Evaluation of the bound is equivalent to solving a stochastic program with simple recourse (Beale 1955, Wets 1966), so computational issues are already well studied.

Our technique works in the other direction, too. By first taking the dual of the recourse LP and then applying the new bounding methodology, we achieve a lower (upper) bound when the primal problem is a minimization (maximization). This result is derived using a restricted dual problem, so it is a *dual* restricted-recourse bound. Because the primal construction yields one bound and the dual construction yields a complementary bound, our techniques give us the important ability to obtain both lower and upper bounds on a problem when other bounds are inapplicable or are more difficult to use.

Like other functional approximations, our method bounds an r-dimensional expectation by a sum of lower-dimensional expectations. For example, if each constraint contains at most one random parameter the primal restricted-recourse bound optimizes a sum of one-dimensional expectations that involve only the random parameters' marginal distributions. This is an advantage because the bounds are much easier to calculate than the original higher dimensional expectation, and the bounds are valid for dependent as well as independent random parameters. That certain dependencies are not captured is a disadvantage because it can weaken the bounds.

The rest of the paper is outlined as follows. In §1, we develop primal restricted-recourse bounds for general, two-stage stochastic-programming problems. We then show how the primal bounding models vary and can be simplified depending on the location of random coefficients in the model, and whether the random coefficients are discretely or continuously distributed. Also, for some simple network examples with stochastic right-hand sides, we show analytically that our bound can be tighter than the complementary Jensen bound. Section 2 develops the dual restricted-recourse bound, shows various bounding models, and analyzes some network examples, all analogous to §1. Section 3 gives computational results for our bounds applied to a semiconductor manufacturing problem. Section 4 gives conclusions and directions for further research.

1. A PRIMAL RESTRICTED-RECOURSE BOUND FOR STOCHASTIC LPs

1.1. Preliminaries

A two-stage stochastic LP with recourse may be written as

SLP-2 $\quad \min_{\mathbf{x} \in X} \mathbf{cx} + Eh(\mathbf{x}, \tilde{\boldsymbol{\xi}}),$

where

$$h(\mathbf{x}, \tilde{\boldsymbol{\xi}}) = \min_{\mathbf{y} \in Y} \tilde{\mathbf{f}} \mathbf{y}$$

$$\text{s.t. } \tilde{D}\mathbf{y} \geqslant \tilde{B}\mathbf{x} + \tilde{\mathbf{d}}.$$

X is a polyhedral set and $\tilde{\boldsymbol{\xi}} = \text{vec}(\tilde{\mathbf{d}}, \tilde{\mathbf{f}}, \tilde{B}\tilde{D})$, where the "vec" operator reads its arguments columnwise to form a single vector. The random vector $\tilde{\boldsymbol{\xi}}$ is defined on a probability space $(\Omega, \mathscr{F}, \mathscr{P})$ with a generic realization denoted ξ. Deterministic linear constrains on $\mathbf{y}$ are represented by $\mathbf{y} \in Y$; these normally include, at least, the nonnegativity restrictions $\mathbf{y} \geqslant \mathbf{0}$.

One motivation for the bounding techniques described in this paper is to aid in the solution of problems of the form of **SLP-2**. However, for ease of presentation, we assume that the first-stage decision vector $\mathbf{x}$ is fixed and focus on bounding the expected value of the redefined recourse function

RF0 $\quad h_0(\tilde{\boldsymbol{\xi}}) = \min_{\mathbf{y} \in Y} \tilde{\mathbf{f}} \mathbf{y}$

$$\text{s.t. } \tilde{D}\mathbf{y} \geqslant \tilde{\mathbf{d}}.$$

We have suppressed the fixed first-stage decision in **RF0**, absorbed the entire right-hand side in $\tilde{\mathbf{d}}$, and redefined $\tilde{\boldsymbol{\xi}}$ to be $\tilde{\boldsymbol{\xi}} = \text{vec}(\tilde{\mathbf{d}}, \tilde{\mathbf{f}}, \tilde{D})$. We will refer to **RF0** as "the second

stage" and call **y** "second-stage variables," but it should be clear that **RF0** also represents stand-alone stochastic LPs such as the stochastic maximum-flow and stochastic PERT problems described in the introduction. We use only "$\geqslant$" constraints in developing our initial bound for **RF0**, but any combination of inequalities and equalities may be used; see §1.2. By convention, matrix D has m rows and n columns, and the remaining matrices and vectors are dimensioned to conform. Throughout this paper we assume that first moment of $\tilde{\xi}$ is finite.

The restricted-recourse bound we develop is based on two results contained in Lemmas 1 and 2 below.

LEMMA 1 (Madansky 1960). *Let* $h(\tilde{\xi}) = \min_{\mathbf{y} \in Y} v(\mathbf{y}, \xi)$ *where* $v(\mathbf{y}, \tilde{\xi})$ *is a real-valued function. Then*

$$Eh(\tilde{\xi}) = E \min_{\mathbf{y} \in Y} v(\mathbf{y}, \tilde{\xi}) \leqslant \min_{\mathbf{y} \in Y} Ev(\mathbf{y}, \tilde{\xi}).$$

LEMMA 2. *Suppose that the following LP has a finite optimal solution:*

LP0 $\quad z_0^* = \min_{\mathbf{y} \geqslant 0} \mathbf{fy}$

$$\text{s.t.} \quad D\mathbf{y} \geqslant \mathbf{d} : \boldsymbol{\pi}$$

$$H\mathbf{y} \geqslant \mathbf{h} : \boldsymbol{\gamma},$$

where $\boldsymbol{\pi}$ *and* $\boldsymbol{\gamma}$ *are dual variables for the indicated constraints. Let* $\boldsymbol{\pi}''$ *be a vector of upper bounds on* $\boldsymbol{\pi}^*$ *for some*

$$(\boldsymbol{\pi}^*, \boldsymbol{\gamma}^*) \in \operatorname*{argmax}_{\substack{\boldsymbol{\pi}D + \boldsymbol{\gamma}H \leqslant \mathbf{f} \\ \boldsymbol{\pi} \geqslant 0, \boldsymbol{\gamma} \geqslant 0}} |\boldsymbol{\pi}\mathbf{d} + \boldsymbol{\gamma}\mathbf{h}|.$$

Also, let

LP1 $\quad z_1^* = \min_{\mathbf{y} \in Y} \mathbf{fy} + \boldsymbol{\pi}''(\mathbf{d} - D\mathbf{y})^+,$

where $Y = \{\mathbf{y} : H\mathbf{y} \geqslant \mathbf{h}, \mathbf{y} \geqslant 0\}$, *and where* $(\mathbf{d} - D\mathbf{y})^+$ *is the positive part of* $(\mathbf{d} - D\mathbf{y})$, *i.e., the vector of constraint violations of* $D\mathbf{y} \geqslant \mathbf{d}$. *Then,* $z_0^* = z_1^*$.

PROOF. The dual of **LP0** is

DP0 $\quad z_0^* = \max_{\boldsymbol{\pi} \geqslant 0, \boldsymbol{\gamma} \geqslant 0} \boldsymbol{\pi}\mathbf{d} + \boldsymbol{\gamma}\mathbf{h}$

$$\text{s.t.} \quad \boldsymbol{\pi}D + \boldsymbol{\gamma}H \leqslant \mathbf{f}.$$

By hypothesis, we may add the upper bounds $\boldsymbol{\pi}'' \geqslant \boldsymbol{\pi}$ without changing the optimal solution value to **DP0**. Therefore, the following LP is equivalent to **DP0**:

DP0a $\quad z_0^* = \max_{\boldsymbol{\pi} \geqslant 0, \boldsymbol{\gamma} \geqslant 0} \boldsymbol{\pi}\mathbf{d} + \boldsymbol{\gamma}\mathbf{h}$

$$\text{s.t.} \quad \boldsymbol{\pi}D + \boldsymbol{\gamma}H \leqslant \mathbf{f}$$
$$\boldsymbol{\pi} \leqslant \boldsymbol{\pi}''.$$

The dual of **DP0a** is **LP1**, so $z_0^* = z_1^*$. (Expand **LP1** into the form of a standard **LP**.) $\square$

Lemma 1 is the "wait-and-see" bound of Madansky (1960) and follows directly from exchanging the order of expectation and optimization. This bound can be applied

to **SLP-2** to obtain a lower bound by allowing first-stage decisions to adapt artificially to every realization of the second stage. This is a relaxation of the original problem and hence yields a lower bound. However, we will apply it to the second-stage problem to obtain an upper bound by insisting that only one second-stage decision be employed for all realizations. This is a restriction and hence leads to an upper bound.

Lemma 2 is related to results on exact penalty functions for nonlinear programs (Zangwill 1967) and aggregation bounds for linear programs (Zipkin 1980). This lemma allows us to restate the primal stochastic LP using an equivalent penalty-based formulation of the problem. Then we apply restricted recourse, Lemma 1, to the reformulation to obtain an upper bound. Bounds on dual variables will sometimes be difficult to obtain, but often will not. For instance, an elastic LP has explicit penalties $\hat{\boldsymbol{\pi}}$ for violating constraints and these penalties are upper bounds on $\boldsymbol{\pi}$.

1.2. The Primal Bound

Here we derive the primal restricted-recourse bound.

THEOREM 1. *Let* $\tilde{\xi} = \text{vec}(\tilde{\mathbf{d}}, \tilde{\mathbf{f}}, \tilde{D})$ *and*

RF0 $\quad h_0(\tilde{\xi}) = \min_{\mathbf{y} \in Y} \tilde{\mathbf{f}}\mathbf{y}$

$$\text{s.t.} \quad \tilde{D}\mathbf{y} \geqslant \tilde{\mathbf{d}} : \boldsymbol{\pi}(\tilde{\xi}).$$

Let $\boldsymbol{\pi}'' \geqslant \boldsymbol{\pi}^*(\tilde{\xi})$, *wp1 (with probability 1), where* $\boldsymbol{\pi}^*(\tilde{\xi})$ *denotes the appropriate subvector of an optimal solution to the dual of* **RF0**. *Then,* $h_{\mathscr{P}}'' \geqslant Eh_0(\tilde{\xi})$ *where*

NLP1 $\quad h_{\mathscr{P}}'' = \min_{\mathbf{y} \in Y} |E\tilde{\mathbf{f}}|\mathbf{y} + \boldsymbol{\pi}''E[(\tilde{\mathbf{d}} - \tilde{D}\mathbf{y})^+].$

PROOF. Let $h_{\mathscr{P}}(\tilde{\xi}) = \min_{\mathbf{y} \in Y} \tilde{\mathbf{f}}\mathbf{y} + \boldsymbol{\pi}''(\tilde{\mathbf{d}} - \tilde{D}\mathbf{y})^+$. Then

$$Eh_0(\tilde{\xi}) = Eh_{\mathscr{P}}(\tilde{\xi}) \qquad\qquad \text{by Lemma 2}$$

$$\leqslant \min_{\mathbf{y} \in Y} E[\tilde{\mathbf{f}}\mathbf{y} + \boldsymbol{\pi}''(\tilde{\mathbf{d}} - \tilde{D}\mathbf{y})^+] \quad \text{by Lemma 1}$$

$$= \min_{\mathbf{y} \in Y} [E\tilde{\mathbf{f}}]\mathbf{y} + \boldsymbol{\pi}''E[(\tilde{\mathbf{d}} - \tilde{D}\mathbf{y})^+]. \quad \square$$

The "$\mathscr{P}$" subscript is used because this is the "primal restricted bound"; the dual restricted bound (§2) will be indicated with a "$\mathscr{D}$" subscript. Of course, we need to describe the linear and nonlinear models that are required to compute the bound proposed in Theorem 1. Before doing that, however, we wish to point out some important features of the preceding result.

REMARK 1. Unlike many other bounds, the primal restricted-recourse bound does not require convexity, concavity, or saddle-function properties of the recourse function $h_0(\xi)$.

REMARK 2. Finding good bounds on optimal dual variables is key to the quality of the proposed bound. Special structures provided by network models, or elastic formulations of

more general models, often enable one to state such bounds a priori. Sections 1.4 and 3 illustrate how these dual bounds may be derived and improved upon.

REMARK 3. If **RF0** is reformulated as an elastic model that includes explicit penalties $\hat{\boldsymbol{\pi}}$ for violating constraints, Theorem 1 follows with $\boldsymbol{\pi}'' = \hat{\boldsymbol{\pi}}$ because $\hat{\boldsymbol{\pi}}$ is an explicit vector of upper bounds on $\boldsymbol{\pi}^*(\tilde{\boldsymbol{\xi}})$. Such explicit penalties also allow Theorem 1 to be generalized to models in which the variables $\mathbf{y}$ are restricted to be integers.

REMARK 4. Theorem 1 adapts to any standard constraint type, not just the "$\geqslant$" constraints of the canonical **LP0**/**RF0** formulation. For example,

$$E \begin{bmatrix} \min_{\mathbf{y} \in Y} & \tilde{\mathbf{f}}\mathbf{y} \\ \text{s.t.} & \tilde{D}\mathbf{y} = \tilde{\mathbf{d}} \end{bmatrix}$$
$$\leqslant \min_{\mathbf{y} \in Y} [E\tilde{\mathbf{f}}]\mathbf{y} + \boldsymbol{\pi}'' E[(\tilde{\mathbf{d}} - \tilde{D}\mathbf{y})^+] - \boldsymbol{\pi}' E[(\tilde{D}\mathbf{y} - \tilde{\mathbf{d}})^+],$$

provided $\boldsymbol{\pi}' \leqslant \boldsymbol{\pi}^*(\tilde{\boldsymbol{\xi}}) \leqslant \boldsymbol{\pi}''$, $wp1$, where $\boldsymbol{\pi}^*(\tilde{\boldsymbol{\xi}})$ is an optimal dual vector associated with constraints $\tilde{D}\mathbf{y} = \tilde{\mathbf{d}}$ for $\tilde{\boldsymbol{\xi}} = \text{vec}(\tilde{\mathbf{d}}, \tilde{\mathbf{f}}, \tilde{D})$.

REMARK 5. **NLP1** is a stochastic program with *simple recourse* (Wets 1966, Ziemba 1974, Wets 1983), which, because of its special structure, can be solved with computational ease compared to stochastic programs with general recourse.

In the next section, we describe the models that must be solved to compute the primal restricted-recourse bound. These models vary depending on whether the random coefficients are continuous or discrete, and become simpler when only certain subsets of these coefficients are, in fact, random.

1.3. Models for Computing the Primal Bound

1.3.1. Discrete Random Variables. When $\tilde{D}$ and $\tilde{\mathbf{d}}$ are discrete random variables that take on a finite number of values—the form of $\tilde{\mathbf{f}}$ is important only through its mean—**NLP1** can be solved as a finite linear program. Let the rth realization of the row vector $(\tilde{D}_{i\cdot}, \tilde{d}_i)$ be denoted $(D_{i\cdot}^r, d_i^r)$, let the set R_i index these realizations, and let $p_i^r = P[(\tilde{D}_{i\cdot}, \tilde{d}_i) = (D_{i\cdot}^r, d_i^r)]$. **NLP1** can now be written as

LP1a
$$\min_{\mathbf{y} \in Y, \mathbf{z} \geqslant 0} [E\tilde{\mathbf{f}}]\mathbf{y} + \sum_{i=1}^m \sum_{r \in R_i} \pi_i'' p_i^r z_i^r$$
$$\text{s.t.} \quad D_{i\cdot}^r \mathbf{y} + z_i^r \geqslant d_i^r, \quad i = 1, \ldots, m, \ r \in R_i.$$

Before looking more closely at **LP1a**, it is important to note that no generality is lost by ignoring first-stage variables in these discrete models. If $\tilde{B}$ of **SLP-2** has discretely distributed random coefficients with a finite number of realizations, the upper-bounding model for **SLP-2** simply

becomes

$$\min_{\mathbf{x} \in X, \mathbf{y} \in Y, \mathbf{z} \geqslant 0} \mathbf{cx} + [E\tilde{\mathbf{f}}]\mathbf{y} + \sum_{i=1}^m \sum_{r \in R_i} \pi_i p_i^r z_i^r$$
$$\text{s.t.} \quad -B_{i\cdot}^r \mathbf{x} + D_{i\cdot}^r \mathbf{y} + z_i^r \geqslant d_i^r, \quad i = 1, \ldots, m, \ r \in R_i,$$

where the rth realization of the vector $(\tilde{D}_{i\cdot}, \tilde{B}_{i\cdot}, \tilde{d}_i)$ is $(D_{i\cdot}^r, B_{i\cdot}^r, d_i^r)$, R_i indexes these realizations, and $p_i^r = P[((\tilde{D}_{i\cdot}, \tilde{B}_{i\cdot}, \tilde{d}_i) = (D_{i\cdot}^r, B_{i\cdot}^r, d_i^r)]$. Some alternative bounds from the literature can be calculated only with fixed first-stage variables $\mathbf{x}$; restricted-recourse bounds could be computed for fixed $\mathbf{x}$, but we see that a tighter bound may be obtained by optimizing the bounding model with respect to $\mathbf{x}$.

Now we consider the size of **LP1a**. For simplicity, assume that Y consists only of nonnegativity constraints. Note that the number of structural constraints in the dual of **LP1a** is n, the dimension of $\mathbf{y}$, and does not depend on the number of realizations. Thus, it may be desirable to solve the dual or apply the dual simplex method (Ziemba 1974, §6). The number of dual decision variables is $\sum_{i=1}^m |R_i|$ and hence can be large. However, the size of this program typically is much smaller than the size of the deterministic equivalent of the corresponding **SLP-2** model because the number of constraints and variables in that problem depends on $\prod_{i=1}^m |R_i|$.

LP1a may be of reasonable size for many models. For example, consider a generalized network with random gains and losses on arc flows such as those used in certain financial models (e.g., Mulvey and Vladimirou 1989). In a model with (a) random returns on asset i from a previous period, (b) a random conversion coefficient for cash to the number of units of asset i, and (c) deterministic exogenous inflow into the associated node, the number of stochastic parameters in $(\tilde{D}_{i\cdot}, \tilde{d}_{i\cdot})$ is two. As a result of this low dimension, $\sum_{i=1}^m |R_i|$ can be modest: If each random coefficient can take on k different values and the original LP has m rows, **LP1a** will have mk^2 rows, whereas the deterministic equivalent would have mk^{2m} second-stage rows.

When the recourse matrix D is deterministic, we have a stochastic program with *fixed recourse* (Wets 1974), and $|R_i|$ depends only on the number of possible outcomes of $\tilde{d}_i$. For this case, **LP1a** simplifies to

LP1b
$$\min_{\mathbf{y} \in Y, \mathbf{z} \geqslant 0} [E\tilde{\mathbf{f}}]\mathbf{y} + \sum_{i=1}^m \sum_{r \in R_i} \pi_i'' p_i^r z_i^r$$
$$\text{s.t.} \quad D_{i\cdot} \mathbf{y} + z_i^r \geqslant d_i^r, \quad i = 1, \ldots, m, \ r \in R_i,$$

where $p_i^r = P[\tilde{d}_i = d_i^r]$.

LP1b can be further simplified. Suppose that each $\tilde{d}_i$ can take on the values $d_i^1 < d_i^2 < \cdots < d_i^{|R_i|}$; for notational simplicity, assume that $D\mathbf{y} \geqslant \mathbf{0}$ for all $\mathbf{y} \in Y$, $d_i^1 > 0$, and

define $d_i^0 = 0$. With $\phi_i^r = P[\tilde{d}_i \geqslant d_i^r]$ **LP1b** simplifies to

LP1c $\quad \min_{\mathbf{y} \in Y, \mathbf{z}} [E\tilde{\mathbf{f}}]\mathbf{y} + \sum_{i=1}^{m} \sum_{r \in R_i} \pi_i'' p_i^r z_i^r$

$\quad$ s.t. $\quad D_{i\cdot}\mathbf{y} + \sum_{r \in R_i} z_i^r \geqslant d_i^{|R_i|}, \quad i = 1, \ldots, m,$

$\quad\quad\quad 0 \leqslant z_i^r \leqslant d_i^r - d_i^{r-1}, \quad i = 1, \ldots, m, \; r \in R_i.$

Therefore, when the right-hand-side $\tilde{\mathbf{d}}$ of the stochastic recourse function consists of discrete random variables, and $\tilde{D}$ is deterministic, our bounding LP, **LP1c**, may be only modestly larger than the "original" deterministic LP, i.e., the stochastic LP with the random vector $\tilde{\mathbf{d}}$ replaced by a deterministic vector $\mathbf{d}$. In particular, the bounding LP will have the same number of structural constraints as the original LP and will include the original structural variables $\mathbf{y}$, but will have an additional set of $\sum_{i=1}^{m} |R_i|$ bounded artificial variables representing possible constraint violations. For solving this problem, Wets (1983) describes an efficient variant of the simplex method that exploits the piecewise-linear convex functions modeled with the terms $\sum_{r \in R_i} z_i^r$. He reports that the solution times for such an algorithm are only "marginally greater" than the times required to solve the same LPs with each summation replaced by a single variable. For example, for an application in finance, Kusy and Ziemba (1986) found the simple-recourse model's solution times to be "... 1.5–2 times that of a related deterministic model ...."

When attempting to solve two-stage programs by utilizing **LP1c**, ordering the realizations of $\tilde{d}_i$ may not be possible if $\tilde{\mathbf{d}}$ depends on $\mathbf{x}$ *and* we attempt to simultaneously optimize over $\mathbf{x}$ and $\mathbf{y}$. However, in some models, second-stage constraints contain random right-hand sides that are separate from those containing first-stage decision variables, and this issue does not arise. This is illustrated in the computational example of §3.

Finally, we note that when both $\tilde{D}$ and $\tilde{\mathbf{d}}$ are deterministic vectors, our bound reduces to the standard Jensen bound for a stochastic LP with a stochastic objective function $\tilde{\mathbf{f}}\mathbf{y}$.

1.3.2 Continuous Random Variables. When $\tilde{D}$ and $\tilde{\mathbf{d}}$ both contain continuous random variables (again $\tilde{\mathbf{f}}$ is unimportant), using Theorem 1 to compute a bound may be difficult because computing expected constraint violations can involve convolutions of random variables whose weights, via the decision variables $\mathbf{y}$, are not known a priori. However, when $\tilde{\mathbf{d}}$ is continuously distributed, and $\tilde{D}$ is deterministic (and in the case of two-stage models, $\tilde{B}$ is deterministic), **NLP1** may be solved efficiently because the objective is just a sum of one-dimensional integrals. For instance, when each $\tilde{d}_i$ is uniformly distributed, the bounding model will be a separable quadratic program. This was shown by Beale (1961) in the context of stochastic programs with simple recourse and is illustrated by a simple example in the next section. Wets (1966) shows that when $\tilde{d}_i$ is an exponential random

variable, the resulting integral has a simple analytical form, too. We will return to this case in an example in §2.3.

Even when the objective is not analytically tractable, **NLP1** can be solved with relative ease when $\tilde{D}$ is deterministic because the objective is a sum of one-dimensional expectations. For example, Kall et al. (1988) solve problems such as **NLP1** by applying sequential-approximation methods with the bounds of Jensen and Edmundson–Madansky. (The dimensional limitations of the E–M bound are not a factor in this combined technique because the bound is applied to one-dimensional expectations.) This method creates a sequence of discrete approximations to the continuous problem, each of which can be formulated and solved as described in §1.3.1. Wets (1983) discusses warm-start procedures for efficiently solving this sequence of approximations.

1.4. An Example

For illustrative purposes, we apply the bound of Theorem 1 to the stochastic maximum-flow (SMF) problem (e.g., Evans 1976). For this problem, $\tilde{D}$ is deterministic and only part of $\tilde{\mathbf{d}}$ is random. Let $G = (N, A)$ be a directed graph where N is the set of nodes and A is the set of arcs. (See Ahuja et al. 1993 for more detailed definitions on network flows.) Arc $k = (i, j)$ is directed from node i to node j, and A includes an artificial "return arc" $k_0 = (t, s)$ that connects the "sink node" t to the "source node" s. Let $FS(i)$ ("forward star of i") and $RS(i)$ ("reverse star of i") denote all arcs directed out of node i and into node i, respectively. Each arc k, except k_0, has a random capacity $\tilde{u}_k \geqslant 0$, which is an upper bound on arc flow y_k; y_{k_0} is unconstrained. We wish to calculate the expected value of the maximum flow through the network between s and t. Restated, the problem is to find $Eh_0(\tilde{\mathbf{u}})$ where

RF0(SMF) $\quad h_0(\tilde{\mathbf{u}}) = \max_{\mathbf{y}} y_{k_0}$

$\quad$ s.t. $\displaystyle\sum_{k \in FS(i)} y_k - \sum_{k \in RS(i)} y_k = 0 \quad \forall i \in N, \quad$ (1)

$\quad\quad\quad\quad\quad\quad\quad y_k \leqslant \tilde{u}_k \quad \forall k \in A - k_0,$

$\quad\quad\quad\quad\quad\quad\quad$ (2)

$\quad\quad\quad\quad\quad\quad\quad y_k \geqslant 0 \quad \forall k \in A - k_0.$

$\quad\quad\quad\quad\quad\quad\quad$ (3)

The function $h_0(\mathbf{u})$ is concave and hence, the Jensen upper bound for the SMF problem is simply $h_0'' = h_0(\bar{\mathbf{u}})$ where $\bar{\mathbf{u}} = E\tilde{\mathbf{u}}$. We now describe our new lower bound, and begin by stating the relevant variant of Theorem 1:

$$E \begin{bmatrix} \max_{\mathbf{y} \in Y} \mathbf{f}\mathbf{y} \\ \text{s.t.} \quad \mathbf{y} \leqslant \tilde{\mathbf{u}} : \pi(\tilde{\mathbf{u}}) \end{bmatrix} = E \left[\max_{\mathbf{y} \in Y} \mathbf{f}\mathbf{y} - \pi''(\mathbf{y} - \tilde{\mathbf{u}})^+ \right]$$

$$\geqslant \max_{\mathbf{y} \in Y} \mathbf{f}\mathbf{y} - \pi'' E[(\mathbf{y} - \tilde{\mathbf{u}})^+],$$

provided $\pi'' \geqslant \pi^*(\tilde{\mathbf{u}})$, $wp1$. Here, $\mathbf{y} \in Y$ represents constraints (1) and (3).

In order to compute the lower bound $h'_{\mathscr{P}}$ for the SMF problem, we require upper bounds on the dual variables for the arc capacity constraints. Since the maximum flow in a network can be increased by at most one unit with the increase in the capacity of an arc by one unit, the dual variable associated with each constraint of type (2) is at most 1. Thus, the following recourse function has the same value as $h_0(\tilde{\mathbf{u}})$ for any realization of $\tilde{\mathbf{u}}$:

$$h_{\mathscr{P}}(\tilde{\mathbf{u}}) = \max_{\mathbf{y} \geqslant \mathbf{0}} \ y_{k_0} - \sum_{k \in A - k_0} (y_k - \tilde{u}_k)^+$$

$$\text{s.t.} \sum_{k \in FS(i)} y_k - \sum_{k \in RS(i)} y_k = 0 \quad \forall i \in N.$$

The lower bound on $Eh_0(\tilde{\mathbf{u}})$ is therefore obtained by solving

NLP1(SMF) $\quad h'_{\mathscr{P}} = \max_{\mathbf{y} \geqslant \mathbf{0}} \ y_{k_0} - \sum_{k \in A - k_0} E[(y_k - \tilde{u}_k)^+]$

$$\text{s.t.} \sum_{k \in FS(i)} y_k - \sum_{k \in RS(i)} y_k = 0 \quad \forall i \in N.$$

If the random variables $\tilde{u}_k$ are discrete, **NLP1(SMF)** becomes a linear program whose objective function is piecewise-linear and concave in $\mathbf{y}$. In §3 we compute bounds for problems with discretely distributed random variables, but at this point it is more revealing to investigate two simple "limiting problems" where the $\tilde{u}_k$ are continuously distributed. We compare the quality of the new lower bound with the Jensen upper bound for these problems.

Let the arc capacities $\tilde{u}_k$ for an SMF problem be uniform random variables on $[0, u''_k]$. Then,

$$E[(y_k - \tilde{u}_k)^+] = \int_0^{y_k} \frac{y_k - u_k}{u''_k} \, du_k = \frac{y_k^2}{2u''_k}.$$

Therefore, computing $h'_{\mathscr{P}}$ will involve solving a separable quadratic programming problem. Next, suppose that G consists of n arcs in parallel between s and t (plus the return arc) and the capacities are i.i.d. uniform random variables on $[0, u'']$. Let $\bar{h}(n)$ denote the true expected maximum flow for n arcs, let $h''_0(n)$ denote the corresponding Jensen upper bound, and let $h'_{\mathscr{P}}(n)$ denote the new lower bound. The Jensen upper bound is exact, i.e., $h''_0(n) = \bar{h}(n) = nu''/2$, because there is only one cut in the network and its average capacity equals the average (expected) maximum flow. The lower bound $h'_{\mathscr{P}}(n)$ can be computed by noting that the flow y_k on each arc k must be the same at optimality, say y, and the total flow will be ny. Thus,

$$h'_{\mathscr{P}}(n) = \max_{y \geqslant 0} \ ny - n\frac{y^2}{2u''} = \frac{nu''}{2},$$

which is exact like the Jensen bound. (Because the recourse function is linear in the arc capacities, exact Jensen and restricted-recourse bounds are actually obtained for any random capacities with finite means; neither the identical nor independent hypothesis is actually required.)

Next, suppose that G consists of the same n arcs placed in series between s and t. The average maximum flow will

be the expected value of the minimum of a set of i.i.d. uniform random variables. This value is $\bar{h}(n) = u''/(n+1)$. In contrast, the Jensen upper bound is $h''_0(n) = u''/2$ so that $\lim_{n \to \infty} h''_0(n)/\bar{h}(n) = \infty$. The lower bound may be obtained by modifying the derivation for the parallel case, noting that the flow on any arc equals the total flow. Thus,

$$h'_{\mathscr{P}}(n) = \max_{y \geqslant 0} \ y - n\frac{y^2}{2u''} = \frac{u''}{2n}.$$

This bound is not tight, but $h'_{\mathscr{P}}(n)/\bar{h}(n) \geqslant 1/2$ for any $n > 0$, and hence it does not become arbitrarily poor for increasing n as does the Jensen bound.

Before leaving this example, we illustrate how the primal restricted-recourse bound can be tightened if better bounds on dual variables are available. The following proposition is a simple extension of Theorem 1.

PROPOSITION 1. *Suppose that bounds on optimal dual variables $\pi^*(\xi)$ are available in* **RF0** *as a function of the parameters ξ, i.e., $\pi''(\tilde{\xi}) \geqslant \pi^*(\tilde{\xi})$, wp1. Then,*

$$Eh_0(\tilde{\xi}) \leqslant \min_{\mathbf{y} \in Y} [E\tilde{\mathbf{f}}]\mathbf{y} + E[\pi''(\tilde{\xi})(\tilde{\mathbf{d}} - \tilde{D}\mathbf{y})^+].$$

For **NLP1(SMF)**, the corresponding functional dual bound is denoted $\pi''(\tilde{\mathbf{u}})$ and the bounding problem can be revised to

NLP2(SMF) $\quad h'_{\mathscr{P}} = \max_{y \geqslant 0} \ y_{k_0} - \sum_{k \in A - k_0} E[\pi''(\tilde{\mathbf{u}})(y_k - \tilde{u}_k)^+]$

$$\text{s.t.} \sum_{k \in FS(i)} y_k - \sum_{k \in RS(i)} y_k = 0 \quad \forall i \in N.$$

We can exploit **NLP2(SMF)** easily in the series-network example by noting that an optimal dual π^*_k in this problem is 1 if $\tilde{u}_k$ is a limiting capacity in the network and is 0 otherwise. So, define

$$\pi''_k(\tilde{\mathbf{u}}) = \begin{cases} 1 & \text{if } \tilde{u}_{k'} \geqslant \tilde{u}_k \ \forall k', \\ 0 & \text{otherwise;} \end{cases}$$

suppose that the $\tilde{u}_k \geqslant 0$ are continuous and i.i.d., and let $\tilde{u}_{\min} = \min_k \tilde{u}_k$. Then, a revised bound for this example is

$$h'_{\mathscr{P}}(n) = \max_{y \geqslant 0} \ y - \sum_{k=1}^{n} E[\pi''_k(\tilde{\mathbf{u}})(y - \tilde{u}_k)^+]$$

$$= \max_{y \geqslant 0} \ y - nE[\pi''_1(\tilde{\mathbf{u}})(y - \tilde{u}_1)^+ | \tilde{u}_1 = \tilde{u}_{\min}] P[\tilde{u}_1 = \tilde{u}_{\min}]$$

$$= \max_{y \geqslant 0} \ y - E[(y - \tilde{u}_{\min})^+]$$

$$= E\tilde{u}_{\min}.$$

This is the exact solution to this problem; we have successfully tightened the restricted-recourse bound. If the random arc capacities have unbounded support, the above optimization must be interpreted as the greatest lower bound (i.e., supremum) instead of a maximization. Similarly, when necessary, the optimization in Proposition 1 and Theorem 1 should be interpreted as the least upper bound, i.e., "min"

is replaced with "inf." (This issue does not arise when Y is compact, which may be assumed, without loss of generality, for finite-valued recourse functions whose random parameters have bounded support.)

In this example, the $\pi''_k(\tilde{\mathbf{u}})$ are optimal dual variables, not just bounds on optimal dual variables. We do not suggest that it will normally be possible to obtain optimal dual variables as a function of the random parameters and thereby find the exact solution to a stochastic programming problem. However, by using dual bounds that are functions of some of the random variables in the problem, it may be possible to improve the quality of a restricted-recourse bound.

2. DUAL RESTRICTED-RECOURSE BOUND

The problem addressed here is to compute a lower bound on $Eh_0(\tilde{\xi})$, where

RF0 $\quad h_0(\tilde{\xi}) = \min_{0 \leqslant \mathbf{y} \leqslant \mathbf{u}} \tilde{\mathbf{f}}\mathbf{y}$

$$\text{s.t.} \quad \tilde{D}\mathbf{y} \geqslant \tilde{\mathbf{d}}.$$

The resulting bound will be complementary to the bound developed in §1. Recall that the primal restricted-recourse bound requires dual bounds $\pi'' \geqslant \pi^*(\tilde{\xi})$, but only for constraints that contain stochastic parameters. Similarly, the dual restricted-recourse bound developed in this section requires simple bounds on $\mathbf{y}$ (or on $\mathbf{y}^*(\tilde{\xi})$, an optimal solution) but only for *columns* of **RF0** that contain stochastic parameters. Analogous to the discussion at the end of §1.4, we may tighten the dual restricted-recourse bound by using bounds on $\mathbf{y}$ of the form $\mathbf{u}(\tilde{\xi})$; this is illustrated in §3. For notational simplicity in this section, we assume that all columns of **RF0** have stochastic parameters and that each y_j has a deterministic simple upper bound, i.e., $y_j \leqslant u_j$ for all j.

2.1. The Dual Bound

THEOREM 2. *Let* $\tilde{\xi} = \text{vec}(\tilde{\mathbf{d}}, \tilde{\mathbf{f}}, \tilde{D})$. *Then,* $h'_{\mathscr{D}} \leqslant Eh_0(\tilde{\xi})$, *where*

NLP2 $\quad \tilde{h}'_{\mathscr{D}} = \max_{\boldsymbol{\pi} \geqslant 0} \boldsymbol{\pi}[E\tilde{\mathbf{d}}] - \sum_{j=1}^{n} u_j E[(\boldsymbol{\pi}\tilde{D}_{\cdot j} - \tilde{f}_j)^+].$

PROOF. By taking the dual of **RF0** and applying Lemma 2, we may re-express $h_0(\tilde{\xi})$ as

$$h_0(\tilde{\xi}) = \max_{\boldsymbol{\pi} \geqslant 0, \, \boldsymbol{\mu} \geqslant 0} \boldsymbol{\pi}\tilde{\mathbf{d}} - \boldsymbol{\mu}\mathbf{u}$$

$$= \max_{\boldsymbol{\pi} \geqslant 0} \boldsymbol{\pi}\tilde{\mathbf{d}} - \sum_{j=1}^{n} u_j(\boldsymbol{\pi}\tilde{D}_{\cdot j} - \tilde{f}_j)^+$$

$$\text{s.t.} \quad \boldsymbol{\pi}\tilde{D} - \boldsymbol{\mu}I \leqslant \tilde{\mathbf{f}}.$$

By applying a restriction, i.e., the "max" variant of Lemma 1, we obtain

$$Eh_0(\tilde{\xi}) \geqslant \max_{\boldsymbol{\pi} \geqslant 0} \boldsymbol{\pi}[E\tilde{\mathbf{d}}] - \sum_{j=1}^{n} u_j E[(\boldsymbol{\pi}\tilde{D}_{\cdot j} - \tilde{f}_j)^+] = h'_{\mathscr{D}}. \quad \square$$

As in §1, convexity of $h_0(\xi)$ is not required to obtain the lower bound in Theorem 2. Naturally, sharp bounds $\mathbf{u}$ on $\mathbf{y}^*(\tilde{\xi})$ will yield tighter lower bounds on $Eh_0(\tilde{\xi})$ and, as with the primal bound, computing $h'_{\mathscr{D}}$ is equivalent to solving a stochastic program with simple recourse. Models that allow efficient computation of the bounds are discussed next.

2.2. Models for Computation

2.2.1. Discrete Random Variables. When $\tilde{D}$ and $\tilde{\mathbf{f}}$ are discrete random variables that take on a finite number of values, **NLP2** can be solved as a linear program. Note that the bound from Theorem 2 depends on $\tilde{\mathbf{d}}$ only through its mean. Let realizations of the jth column vector $(\tilde{f}_j, \tilde{D}_{\cdot j})$ be denoted $(f_j^r, D_{\cdot j}^r)$ where $r \in R_j$, and let the corresponding probability mass function be denoted $P[(\tilde{f}_j, \tilde{D}_{\cdot j}) = (f_j^r, D_{\cdot j}^r)] = p_j^r$. With this notation, **NLP2** may be written as

$$\max_{\boldsymbol{\pi} \geqslant 0, \, \mathbf{z} \geqslant 0} \boldsymbol{\pi}[E\tilde{\mathbf{d}}] - \sum_{j=1}^{n} \sum_{r \in R_j} p_j^r u_j z_j^r \tag{4}$$

$$\text{s.t.} \quad \boldsymbol{\pi}D_{\cdot j}^r - z_j^r \leqslant f_j^r, \quad j = 1, \ldots, n, \, r \in R_j.$$

Taking the dual of (4) yields the following primal formulation:

$$\max_{\mathbf{y} \geqslant 0} \quad \sum_{j=1}^{n} \sum_{r \in R_j} f_j^r y_j^r$$

$$\text{s.t.} \quad \sum_{j=1}^{n} \sum_{r \in R_j} D_{\cdot j}^r y_j^r \geqslant E\tilde{\mathbf{d}}, \tag{5}$$

$$y_j^r \leqslant p_j^r u_j, \quad j = 1, \ldots, n, \, r \in R_j.$$

In the case of a two-stage stochastic LP, $E\tilde{\mathbf{d}}$ would simply be replaced by $E\tilde{\mathbf{d}} + [E\tilde{B}]\mathbf{x}$, and the minimization would be taken over $\mathbf{x}$ and $\mathbf{y}$.

The number of structural constraints in (5) does not depend on the number of realizations, and the number of decision variables is $\sum_{j=1}^{n} |R_j|$. While this value may be large, it is typically much smaller than $\prod_{j=1}^{n} |R_j|$, which determines the size of the deterministic equivalent for the original problem **SLP-2**. When the number of stochastic parameters per column is modest, computing the bound via (5) will be practical. For example, if the recourse problem involves optimization over a typical generalized network, $|R_j|$ would be at most 2, corresponding to stochastic cost and gain (or loss) coefficients.

When $\tilde{D}$ is deterministic, the size of the model depends only on the number of realizations in the marginal distributions of $\tilde{f}_j$. The form of the bounding program (5) does not change, except that $D_{\cdot j}^r$ becomes $D_{\cdot j}$. When both $\tilde{D}$ and $\tilde{\mathbf{f}}$ are deterministic, it is clear from (5) that the proposed lower bound is equivalent to Jensen's bound for a stochastic LP with a random right-hand side.

2.2.2. Continuous Random Variables. When $\tilde{D}$ is deterministic but $\tilde{\mathbf{f}}$ is continuous, computing the lower bound directly

from the **NLP2** formulation is possible because it involves only one-dimensional integrals. The remarks from §1.3.2 regarding analytically available nonlinear objective functions and the use of bounding and approximation techniques are also relevant here. In the next section, we give an example where the random cost coefficients are exponential random variables.

2.3. Example

A simple example to which the above technique may be applied is the stochastic shortest-path (SSP) problem (e.g., Frank 1969, Mirchandani 1976). We are interested in computing the expected length of a shortest path between nodes s and t in a directed network G where the length (e.g., time, distance, or cost) of each arc $k = (i, j)$ is a random variable $\tilde{c}_k$. Assuming that there are no negative-length cycles in G, the deterministic shortest-path problem may be formulated as a network flow problem that routes one unit of flow as cheaply as possible from s to t. Thus, SSP computes $Eh_0(\tilde{\mathbf{c}})$, where $h_0(\tilde{\mathbf{c}})$ is defined by

$$\textbf{RF0(SSP)} \quad h_0(\tilde{\mathbf{c}}) = \min_{\mathbf{y} \geqslant \mathbf{0}} \sum_{k \in A} \tilde{c}_k y_k$$

$$\text{s.t.} \sum_{k \in FS(i)} y_k - \sum_{k \in RS(i)} y_k$$

$$= \begin{cases} 1 & \text{for } i = s, \\ 0 & \forall i \in N - \{s, t\}, \\ -1 & \text{for } i = t. \end{cases}$$

Applying Jensen's inequality to this problem yields an upper bound $h_0'' = h_0(\tilde{\mathbf{c}})$. In order to derive our complementary lower bound we take the dual of the shortest path problem:

$$\textbf{RF0(SSP, dual)} \quad h_0(\tilde{\mathbf{c}}) = \max_{\boldsymbol{\pi}} \pi_s - \pi_t$$
$$\text{s.t.} \quad \pi_i - \pi_j \leqslant \tilde{c}_k$$
$$\forall k = (i, j) \in A.$$

Using the fact that $y_k \leqslant 1$, the bound may be stated:

$$h_{\mathcal{D}}' = \max_{\boldsymbol{\pi}} \quad (\pi_s - \pi_t) - \sum_{k=(i,j) \in A} E(\pi_i - \pi_j - \tilde{c}_k)^+. \quad (6)$$

We illustrate this bound with an example in which the nonlinear terms in the objective of (6) may be integrated analytically. Suppose that each arc k has length $\tilde{c}_k$ that is distributed as an exponential random variable with rate λ_k, i.e., $F_k(c) = 1 - e^{-\lambda_k c}$ for $c \geqslant 0$. In this case, the nonlinear objective terms become

$$\sum_{k=(i,j) \in A} \int_0^{\pi_i - \pi_j} (\pi_i - \pi_j - c) \lambda_k e^{\lambda_k c} \, dc$$

$$= \sum_{k=(i,j) \in A} \left[(\pi_i - \pi_j) + \frac{1}{\lambda_k} \exp[-\lambda_k(\pi_i - \pi_j)] - \frac{1}{\lambda_k} \right].$$

$$(7)$$

For an SSP with exponential arc lengths, we may bound the optimal objective from below by solving the unconstrained optimization problem given by (6) and (7). We investigate two limiting cases here to compare the accuracy of the new lower bound to the complementary Jensen bound.

Suppose that G consists of n arcs, arranged in parallel, with lengths that are i.i.d. exponential random variables with common rate λ. Let $\bar{h}(n)$ denote this network's expected shortest path length. This is just the expected value of the minimum of n i.i.d. exponential random variables, so $\bar{h}(n) = 1/n\lambda$. The Jensen upper bound, denoted $h_0''(n)$, is the expected length of any arc, so $h_0''(n) = 1/\lambda$ and $\lim_{n \to \infty} h_0''(n)/\bar{h}(n) = \infty$; i.e., the bound becomes arbitrarily poor as n increases.

The dual restricted-recourse bound for this problem, denoted $h_{\mathcal{D}}'(n)$, is better. Defining $y = \pi_s - \pi_t$, the objective of the bounding problem may be written as

$$y - n \left[y + \frac{1}{\lambda} e^{-\lambda y} - \frac{1}{\lambda} \right]. \quad (8)$$

Minimization of this function through calculus yields

$$h_{\mathcal{D}}'(n) = \frac{1}{\lambda} + \frac{n-1}{\lambda} \log \left(1 - \frac{1}{n} \right).$$

It is easy to show that $\lim_{n \to \infty} h_{\mathcal{D}}'(n)/\bar{h}(n) = 1/2$. Thus the bound is not tight, but it does not become worthless for large values of n. The structure of the dual formulation of this problem is identical to the series SMF problem of §1.4 so that the same technique may be used to obtain an exact bound. In particular, this result is achieved using the functional primal bounds

$$u_k(\tilde{\mathbf{c}}) = \begin{cases} 1 & \text{if } \tilde{c}_{k'} \geqslant \tilde{c}_k \ \forall k', \\ 0 & \text{otherwise.} \end{cases}$$

Finally, suppose that G consists of n arcs in series where the arc lengths are random variables. The true expected shortest path length is the expectation of a sum of n random variables, so $\bar{h}(n) = E \sum_{k \in A} \tilde{c}_k = n/\lambda$. The Jensen upper bound is just the sum of the individual expectations, so it is exact.

To compute our bound, we begin with the dual formulation. Because of the series structure of the network, the inequalities in the dual must hold with equality. With $y = \pi_s - \pi_t$, the dual recourse function reduces to

$$\max_y y$$
$$\text{s.t.} \quad y \leqslant \sum_{k \in A} \tilde{c}_k.$$

This leads to the bound $h_{\mathcal{D}}'(n) = E \sum_{k \in A} \tilde{c}_k = n/\lambda$, and thus the restricted-recourse lower bound is also exact.

3. A COMPUTATIONAL EXAMPLE

This section describes a pilot model for capacity expansion in the semiconductor manufacturing industry and then uses

the model to demonstrate the formulation and computation of restricted-recourse bounds. This model is an extension of Stafford (1997).

In this strategic model, budget-constrained purchases of semiconductor wafer-fabrication machines are made in the first stage so that uncertain demands for wafers of various types and quantities can be met in the second stage as best possible. Demand distributions cover demands over a one-month period, starting about one year in the future—one year being the approximate lead time for delivery and set-up of the fabrication machines. Because of this long lead time, there is significant uncertainty in what the actual demands will be.

Production of a wafer requires multiple steps in a specified order, although a single machine may be able to perform more than one step. Because the model is strategic, however, it does not schedule wafer production per se. Rather, it determines production quantities, ensures that the total number of steps of each type is sufficient to meet production quantities, and ensures that the total time required to perform the assigned steps does not exceed available machine time.

Because demands are aggregates over the period of a month, and because there is a wealth of experience with most of the machine types, the model assumes that the availability of the machines is deterministic. That is, the number of actual production hours available on each machine (accounting for down-time for maintenance and repair) is assumed known. On the other hand, because some of the wafer types that the machines will be producing are new, there is uncertainty in the processing time that the steps will require. Therefore, the second stage of the model incorporates uncertainty in production rates as well as demands.

3.1. The Model

The two-stage stochastic mixed-integer program (MIP) for wafer production (WMIP) is specified below.

Indices and Index Sets

$m \in M$	Machine types
$w \in W$	Wafer (product) type
$s \in S$	Process steps
$S_w \subseteq S$	Process steps required by wafer type w
$S_{wm} \subseteq S_w$	Process steps for wafer w that can be performed on machine m
$W_m \subseteq W$	Wafers that may be processed on machine m
$M_{sw} \subseteq M$	Machines that can perform step s for wafer w

Data

u_m	Number of existing machines of type m
c_m	Unit cost to purchase a machine of type m
B	Total budget for purchasing machines
T_m	Production time available per machine of type m
ρ_m	Unit shortage penalty for wafer type w
$\tilde{t}_{swm}$	Random processing time for step s for w on m
$\tilde{d}_w$	Random demand for wafer type w

Decision Variables

x_m	Number of machines of type m to purchase (first stage)
y_w	Number of wafers of type w to produce (second stage)
z_{swm}	Number of steps s for wafer w to assign to machine m (second stage)
v_w	Unmet demand for wafer type w (second stage)

Formulation

$$\textbf{WMIP} \quad z^* = \min_{\mathbf{x}} \; Ef(\mathbf{x}, \tilde{\mathbf{d}}, \tilde{\mathbf{t}})$$

$$\text{s.t.} \sum_{m \in M} c_m x_m \leqslant B \tag{9}$$

$$x_m \in Z_+ \quad \forall m \in M,$$

$$\text{where} \quad f(\mathbf{x}, \tilde{\mathbf{d}}, \tilde{\mathbf{t}}) = \min_{\mathbf{v}, \mathbf{y}, \mathbf{z}} \sum_{w \in W} \rho_m v_m,$$

$$\text{s.t.} \sum_{w \in W_m} \sum_{s \in S_{wm}} \tilde{t}_{swm} z_{swm} \leqslant T_m(u_m + x_m) \quad \forall m \in M, \tag{10}$$

$$\sum_{m \in M_{sm}} z_{swm} - y_w \geqslant 0 \quad \forall w \in W, \, s \in S_w, \tag{11}$$

$$y_w + v_w \geqslant \tilde{d}_w \quad \forall w \in W, \tag{12}$$

$$y_w, \, v_w \geqslant 0 \quad \forall w \in W,$$

$$z_{swm} \geqslant 0 \quad \forall w \in W, \, s \in S_w, \, m \in M_{sw}.$$

Constraint (9) is the first-stage budget constraint. Second-stage constraints limit the processing time on machines to time available (10), ensure that processing steps are assigned to meet production quantities (11), and meet demand elastically (12). First-stage variables are required to be integral because the number of machines to be purchased is small, about six. Second-stage variables are continuous because the number of wafers being produced is large, in the hundreds. We note that **WMIP** is similar in structure to a production model developed by Higle and Sen (1994), except that their model does not involve separate process steps on different machines.

3.2. Restricted-Recourse Bounds for WMIP

For the purpose of stating the restricted-recourse bounding models, we need additional definitions and assumptions:

1. Demands, $\tilde{d}_w$, are discretely distributed with realizations d_w^l, for $l \in L_w = \{1, 2, \ldots, |L_w|\}$, $d_w^0 \equiv 0, d_w^{l-1} < d_w^l$ for $l \in L_w$, and $\phi_w^l \equiv P[\tilde{d}_w \geqslant d_w^l]$ for $l \in L_w$. In the primal bounding model, unmet demand v_w is expanded by demand level l to v_w^l.

2. Because ρ_w is an explicit elastic penalty on the demand constraint (12) indexed by w, ρ_w is an upper bound on the dual variable associated with that constraint.

3. Processing times $\tilde{t}_{swm} > 0$ are discretely distributed with realizations $t_{swm}^i \in I_{swm}$, and mass functions $p_{swm}^i \equiv P[\tilde{t}_{swm} = t_{swm}^i]$, $i \in I_{swm}$. In the dual bounding model, assignment variables z_{swm} are expanded to z_{swm}^i, $i \in I_{swm}$.

For the primal bounding model, we require $R_m = \prod_{w \in W_m, \, s \in S_{wm}} I_{swm}$, the sample space for the vector of processing times that corresponds to each machine. Let $p_m^r = P[\tilde{t}_{swm} = t_{swm}^r \; \forall w \in W_m, \, s \in S_w]$ for $r \in R_m$. In the primal bounding model, assignment variables z_{swm} are expanded to z_{swm}^r, $r \in R_m$, and constraint-violation variables for the expanded versions of constraint (10) are denoted g_m^r.

4. The dual variables associated with (10) are nonpositive, and thus lower bounds on these variables are needed for the primal restricted-recourse model (see Remark 4 in §1.2). These bounds can be made functions of the realization r (see §1.4). We use $\pi_m^{\prime r} = \min_{w \in W_m, \, s \in S_{wm}} \{-\rho_w / t_{swm}^r\}$.

5. Upper bounds on primal variables z_{swm} are needed to compute the dual restricted-recourse lower bound. These bounds can be made functions of both $T_m(u_m + x_m)$ and the realization of $\tilde{t}_{swm}$ and yield constraints $z_{swm}^i \leqslant T_m(u_m + x_m)/t_{swm}^i$.

6. The restricted-recourse bounds are valid for random parameters with general dependency structures. With respect to both the primal and dual bounding models, we require only the marginal distributions of the demand parameters. However, to calculate the primal bound we must compute p_m^r, and hence we must specify the structure relating the processing times on each machine. We will assume that for each m, $\tilde{t}_{swm}$, $\forall w \in W_m$, $s \in S_w$ are independent.

The model for the primal restricted-recourse bound, with first-stage solution denoted $\mathbf{x}_{\mathscr{P}}^*$, is

$$\textbf{WMIP-}\mathscr{P} \; \min_{\mathbf{x}, \mathbf{g}, \mathbf{v}, \mathbf{y}, \mathbf{z}} \sum_{w \in W} \sum_{l \in L_w} \rho_w \phi_w^l v_w^l - \sum_{m \in M} \sum_{r \in R_m} p_m^r \pi_m^{\prime r} g_m^r$$

$$\text{s.t.} \quad \sum_{w \in W_m} \sum_{s \in S_{wm}} t_{swm}^r z_{swm} - g_m^r$$

$$\leqslant T_m(u_m + x_m) \quad \forall m \in M, \, r \in R_m,$$

$$\sum_{m \in M_{sw}} z_{swm} - y_w \geqslant 0 \quad \forall w \in W, \, s \in S_w,$$

$$y_w + \sum_{l \in L_w} v_w^l \geqslant d_w^{|L_w|} \quad \forall w \in W,$$

$$\sum_{m \in M} c_m x_m \leqslant B,$$

$$x_m \in Z_+ \quad \forall m \in M,$$

$$y_w \geqslant 0 \quad \forall w \in W,$$

$$0 \leqslant v_w^l \leqslant d_w^l - d_w^{l-1} \quad \forall w \in W, \, l \in L_w,$$

$$g_m^r \geqslant 0 \quad \forall m \in M, \, r \in R_m,$$

$$z_{swm} \geqslant 0 \quad \forall w \in W, \, s \in S_w, \, m \in M_{sw}.$$

The model for the dual restricted-recourse bound, with first-stage solution denoted $\mathbf{x}_{\mathscr{D}}^*$, is

$$\textbf{WMIP-}\mathscr{D} \; \min_{\mathbf{x}, \mathbf{v}, \mathbf{y}, \mathbf{z}} \sum_{w \in W} \rho_w v_w$$

$$\text{s.t.} \quad \sum_{w \in W_m} \sum_{s \in S_{wm}} \sum_{i \in I_{swm}} t_{swm}^i z_{swm}^i$$

$$\leqslant T_m(u_m + x_m) \quad \forall m \in M,$$

$$\sum_{m \in M_{sw}} \sum_{i \in I_{swm}} z_{swm}^i - y_w \geqslant 0$$

$$\forall w \in W, \, s \in S_w,$$

$$y_w + v_w \geqslant E\tilde{d}_w \quad \forall w \in W,$$

$$\sum_{m \in M} c_m x_m \leqslant B,$$

$$x_m \in Z_+ \quad \forall m \in M,$$

$$y_w, v_w \geqslant 0 \quad \forall w \in W,$$

$$0 \leqslant z_{swm}^i \leqslant p_{swm}^i T_m(u_m + x_m)/t_{swm}^i$$

$$\forall m \in M, \, w \in W_m,$$

$$s \in S_{wm}, \, i \in I_{swm}.$$

3.3. Computational Results

Here we test the accuracy and speed of computation for the bounds produced by **WMIP-**$\mathscr{P}$ and **WMIP-**$\mathscr{D}$. We also examine the tightness of the bounds as the variability of the random processing times increases.

The test scenario has 27 machine types (only a subset of all types will be purchased), 10 wafer types, up to 7 productions steps for each wafer, 6 existing machines, each of a different type, and an unmet demand penalty of $\rho_w = 1$ for all wafer types w. The budget B and costs c_m allow for the purchase of six machines. For all wafer types w, demands have realizations $d_w^l \in \{50 + 25l, \; l = 1, \ldots, 5\}$ (monthly data have been scaled to units for a single day), with mass function $P[\tilde{d}_w = d_w^l] \equiv p_w^l = 0.1, 0.2 .0.4, 0.2, 0.1$ for $l = 1, \ldots, 5$, respectively. Available time is $T_m = 800$ minutes for each machine. Mean processing times $E\tilde{t}_{swm}$ range between 2 and 9 minutes, and $\tilde{t}_{swm}$ is a binary random variable with equal probability of taking on either value. Specifically, the realizations are $(1 - \alpha)E\tilde{t}_{swm}$ and $(1 + \alpha)E\tilde{t}_{swm}$, where α is a parameter that will be varied to test sensitivity of the bounds to the variability of these random parameters. Our data and formulation limit us to handling two or three realizations of each processing time because of the exponential expansion required in the structural constraints indexed by R_m in the primal bounding model. **WMIP** has a total of $|W| + \sum_{w \in W} |S_w| = 10 + 52 = 62$ random parameters, and under the assumption that the random parameters are independent (one probability model for which the bounds are valid), has $5^{10} \cdot 2^{52}$ total scenarios.

All testing is performed on an IBM RS-6000 Model 590 computer with 512 megabytes of RAM. Models are generated using GAMS (Brooke et al. 1992) and solved using the OSL (1993) mixed-integer solver. **WMIP-**$\mathscr{P}$ has 1152 variables and 1292 structural constraints while **WMIP-**$\mathscr{D}$ has 325 variables and 282 structural constraints. All MIPS are solved to within an absolute tolerance of 0.1 units, which is accounted for in the reported results.

Table 1 lists computational results with six different values of α. The table lists the following values:

1. The value of α.

Table 1. Computational results for **WMIP**-$\mathscr{P}$ and **WMIP**-$\mathscr{D}$.

	Lower bounds		Upper bounds			Solution Secs.		
α	**WLP**-$\mathscr{D}$	**WMIP**-$\mathscr{D}$	**WMIP**-$\mathscr{P}$	**WMIPf**-$\mathscr{P}$	MC	**WMIP**-$\mathscr{D}$	**WMIP**-$\mathscr{P}$	**WMIPf**-$\mathscr{P}$
0.00	114.63	168.11	233.99	233.99	179.27(4.69)	8.19	46.95	2.10
0.05	76.33	127.51	263.28	263.28	173.30(4.70)	6.09	62.40	3.33
0.10	34.24	86.57	292.43	292.43	171.35(4.39)	6.38	58.98	3.58
0.15	0.00	43.88	320.73	320.73	162.68(4.42)	2.08	60.61	3.75
0.20	0.00	0.00	346.95	376.29	158.64(4.41)	0.53	56.56	3.09
0.25	0.00	0.00	371.99	380.76	152.67(4.86)	0.20	47.13	2.91

WMIP-$\mathscr{D}$ Dual lower-bounding model.
WMIP-$\mathscr{P}$ Primal upper-bounding model.
WMIPf-$\mathscr{P}$ Primal upper-bounding model with **x** fixed to $\mathbf{x}_{\mathscr{D}}^*$.
WLP-$\mathscr{D}$ LP relaxation of **WMIP**-$\mathscr{D}$.
MC   Monte Carlo model: upper bound (std. dev.) for **WMIP** with $\mathbf{x} = \mathbf{x}_{\mathscr{P}}^*$

2. The lower bound provided by the LP relaxation of **WMIP**-$\mathscr{D}$. (A valid lower bound for **WMIP** is obtained by solving any relaxation of **WMIP**-$\mathscr{D}$.)

3. The lower bound provided by solving **WMIP**-$\mathscr{D}$ (with integer restrictions).

4. The upper bound provided by solving **WMIP**-$\mathscr{P}$.

5. The upper bound provided by solving **WMIP**-$\mathscr{P}$ with **x** fixed at $\mathbf{x}_{\mathscr{D}}^*$, the solution derived from **WMIP**-$\mathscr{D}$. (A valid upper bound on z^* is obtained by solving any restriction of **WMIP**-$\mathscr{D}$.)

6. A crude Monte Carlo estimate of an upper bound on z^*, with **x** fixed at $\mathbf{x}_{\mathscr{P}}^*$. This is computed by solving the second-stage recourse problem from **WMIP** (with $\mathbf{x} = \mathbf{x}_{\mathscr{P}}^*$) for 200 independent and identically distributed observations of the vector of random parameters. For this calculation we assume that the entire collection of random demands and processing times are mutually independent.

7. The solution time for the dual bounding model **WMIP**-$\mathscr{D}$. (We do not make the obvious simplifications of either model when $\alpha = 0$, i.e., when the $\tilde{t}_{swm}$ are constants.)

8. The solution time for the primal bounding model **WMIP**-$\mathscr{P}$.

9. The solution time for the primal bounding model **WMIP**-$\mathscr{P}$ with **x** fixed at $\mathbf{x}_{\mathscr{D}}^*$.

Table 1 demonstrates a number of issues regarding this application of restricted-recourse bounds. Both the primal upper-bounding model **WMIP**-$\mathscr{P}$ and the dual lower-bounding model **WMIP**-$\mathscr{D}$ are relatively easy to solve; see the first two columns of solution times. Of the two models, **WMIP**-$\mathscr{P}$ is larger and more difficult to solve because of the constraints indexed by R_m.

We note that the solution to the LP relaxation of **WMIP**-$\mathscr{D}$ does not provide a very tight bound on the optimal solution value (see the lower bounds from **WLP**-$\mathscr{D}$ and **WMIP**-$\mathscr{D}$), so we cannot expect to use this easier-to-compute lower bound directly. This also means that optimally solving **WMIP**-$\mathscr{P}$ or **WMIP**-$\mathscr{D}$ with a branch-and-bound algorithm is likely to become more difficult as

the size of the feasible set of first-stage solutions increases. For instance, when the budget allows the purchase of nine machines and $\sum_m u_m = 3$, solution times increase by a factor of about 200%.

Rather than solving **WMIP**-$\mathscr{P}$ exactly to compute an upper bound on z^*, a bound can be obtained by solving a restriction of **WMIP**-$\mathscr{P}$. Fixing **x** to an estimate of $\mathbf{x}^*$ yields an easier-to-solve LP. One obvious estimate of $\mathbf{x}^*$ is $\mathbf{x}_{\mathscr{D}}^*$ because it is feasible and relatively easy to compute. The upper bounds obtained using this estimate are listed under **WMIPf**-$\mathscr{P}$ in the table. For our test scenario, $\mathbf{x}_{\mathscr{D}}^*$ is an optimal solution to **WMIP**-$\mathscr{P}$ for $\alpha = 0$, 0.05, 0.10, 0.15. The lower bounding model is not very informative for $\alpha = 0.20$ and $\alpha = 0.25$ and this hints that $\mathbf{x}_{\mathscr{D}}^*$ may not be a very good estimate of $\mathbf{x}^*$. In fact, the bound derived from **WMIPf**-$\mathscr{P}$ does deteriorate somewhat relative to **WMIP**-$\mathscr{P}$ for these values of α.

Clearly, the primal and dual bounding models do not yield sufficiently tight bounds to be useful as a stand-alone solution technique for **WMIP**. However, this is understandable. First, z^* is a relatively high-dimensional expectation (with 62 random parameters), and it would be surprising if a deterministic bounding technique yielded tight bounds. Second, except for the machine-wise independence assumption on $\tilde{t}_{swm}$, $\forall w \in W_m$, $s \in S_w$, the numerical values from Table 1 are valid for a wide range of dependency structures on the random parameters; because of this, we cannot expect particularly tight bounds. Furthermore, we envision the primal and dual bounds being applied like other bounds within sequential-approximation procedures. The gaps in the bounds of Table 1 are reasonably consistent with the initial gaps—prior to refining the bounds by applying them in a conditional fashion—found in the computational work of Edirisinghe and You (1996), Edirisinghe and Ziemba (1996), and Frauendorfer (1992).

4. CONCLUSIONS

This paper has developed primal and dual versions of a restricted-recourse bound for bounding the expected value

of a stochastic recourse function that evaluates a linear program with random coefficients. We show that any such recourse function can be written as an equivalent function with random coefficients appearing only in the objective function of a primal or dual reformulation of the problem. Taking the expectation of the reformulation and then exchanging the order of expectation and optimization yields the bound.

When only the objective function or right-hand side of the original recourse function is random, one of our bounds is equivalent to the Jensen bound, and our other bound is complementary. So, restricted recourse provides an alternative to standard bounds (e.g., Edmundson–Madansky, piecewise-linear) for obtaining a complementary bound. However, restricted-recourse bounds can be applied when the Jensen bounds and other bounds that exploit convexity cannot be employed (or become computationaly prohibitive). Such is the case when random coefficients appear in some combination of the right-hand side, objective function, and constraint matrix. Significantly, restricted recourse can be used to obtain both lower and upper bounds for such problems by operating on both primal and dual reformulations of the recourse problem.

Computing a restricted-recourse bound is equivalent to solving a stochastic program with simple recourse. For continuous distributions of the random coefficients, the bounding models are convex nonlinear programming problems. When distributions are discrete, the models are LPs with piecewise-linear objective functions. Limiting examples show that the new bounds can be tighter than complementary Jensen bounds, and applying the bounds to a semiconductor manufacturing model demonstrates that the bounds can be computed with modest effort.

There are many avenues open for utilizing, refining, and extending the techniques described in this paper. We have already successfully employed the bounds in a sequential-approximation algorithm for stochastic network-interdiction problems, but further work is necessary to solve larger and more general problems. It is clear that primal restricted-recourse bounds can be used with recourse functions modeled as elastic integer programs, but the quality and computational efficiency of such bounds remains untested. Also, additional research will be needed in order to more effectively employ our bounds for general (non-network) LP recourse functions: Optimal primal and dual variables must be bounded to use restricted recourse, and good bounds for general models may be difficult to obtain.

ACKNOWLEDGEMENTS

REFERENCES

Ahuja, R.K., T.L. Magnanti, J.B. Orlin. 1993. *Network Flows*. Prentice Hall, Englewood Cliffs, NJ.

Beale, E.M.L. 1955. On minimizing a convex function subject to linear inequalities. *J. Royal Statist. Soc.* **17B** 173–184.

——. 1961. The use of quadratic programming in stochastic linear programming. P-2404, The RAND Corporation, Santa Monica, CA.

Birge, J.R. 1985. Aggregation bounds in stochastic linear programming. *Math. Programming* **31** 25–41.

——, J.H. Dulá. 1991. Bounding separable recourse functions with limited distribution information. *Ann. Oper. Res.* **30** 277–298.

——, S.W. Wallace. 1988. A separable piecewise linear upper bound for stochastic linear programs. *SIAM J. Control and Optimiz.* **26** 725–739.

——, R.J.-B Wets. 1986. Designing approximation schemes for stochastic optimization problems, in particular, for stochastic programs with recourse. *Math. Programming Study* **27** 54–102.

——, ——. 1987. Computing bounds for stochastic programming problems by means of a generalized moment problem. *Math. Oper. Res.* **12** 149–162.

——, ——. 1989. Sublinear upper bounds for stochastic programs with recourse. *Math. Programming* **43** 131–149.

Brooke, A., D. Kendrick, A. Meeraus. 1992. *GAMS, A User's Guide, Release 2.25*. Scientific Press, San Francisco, CA.

Cariño, D.R., T. Kent, D.H. Meyers, C. Stacy, M. Sylvanus, A.L. Turner, K. Watanabe, W.T. Ziemba. 1994. The Russell-Yasuda Kasia model: an asset/liability model for a Japanese insurance company using multistage stochastic programming. *Interfaces* **24** 29–49.

Cormican, K.J., D.P. Morton, R.K. Wood. 1998. Solving a stochastic network interdiction problem. *Oper. Res.* **46** 184–197.

Dantzig, G.B. 1995. Linear programming under uncertainty. *Management Sci.* **1** 197–206.

——, P.W. Glynn. 1990. Parallel processors for planning under uncertainty. *Ann. Oper. Res.* **22** 1–21.

——, P.W. Glynn, M. Avriel, J.C. Stone, R. Entriken, M. Nakayama. 1989. Decomposition techniques for multi-area generation and transmission planning under uncertainty. EPRI Report 2940-1.

Dulá, J.H. 1992. An upper bound on the expectation of simplicial functions of multivariate random variables. *Math. Programming* **55** 69–80.

Dupačová, J. 1994. Applications of stochastic programming under incomplete information. *J. Computational and Appl. Math.* **56** 113–125.

Edirisinghe, N.C.P. 1996. New second-order bounds on the expectation of saddle functions with applications to stochastic linear programming. *Oper. Res.* **44** 909–922.

——, G-M. You. 1996. Second-order scenario approximation and refinement in optimization under uncertainty. *Ann. Oper. Res.* **64** 143–178.

W.T. Ziemba 1992. Tight bounds for stochastic convex programs. *Oper. Res.* **40** 660–677.

——, ——. 1994a. Bounds for two-stage stochastic programs with fixed recourse. *Math. Oper. Res.* **19** 292–313.

——, ——. 1994b. Bounding the expectation of a saddle function with application to stochastic programming. *Math. Oper. Res.* **19** 314–340.

——, ——. 1996. Implementing bounds-based approximation in convex–concave two-stage stochastic programming. *Math. Programming* **75** 295–325.

Edmundson, H.P. 1956. Bounds on the expectation of a convex function of a random variable. P-982, The RAND Corporation, Santa Monica, CA.

Ermoliev, Y. 1983. Stochastic quasigradient methods and their applications to systems optimization. *Stochastics* **9** 1–36.

Evans, J.R. 1976. Maximum flow in a probabilistic graph—the discrete case. *Networks* **6** 161–183.

Frank, H. 1969. Shortest paths in probabilistic networks. *Oper. Res.* **17** 583–599.

Frauendorfer, K. 1988. Solving SLP recourse problems with arbitrary multivariate distributions—the dependent case. *Math. Oper. Res.* **13** 377–394.

——. 1992. Stochastic two-stage programming. *Lecture Notes in Economics and Mathematical Systems 392*, Springer-Verlag, Berlin.

Fulkerson, D.R. 1962. Expected critical path lengths in PERT networks. *Oper. Res.* **10** 808–817.

Gassmann, H.I. 1989. Optimal harvest of a forest in the presence of uncertainty. *Canadian J. Forest Res.* **19** 1267–1274.

——, W.T. Ziemba. 1986. A tight upper bound for the expectation of a convex function of a multivariate random variable. *Math. Programming Study* **27** 39–53.

Higle, J.L., S. Sen. 1991. Stochastic decomposition: an algorithm for two-stage linear programs with recourse. *Math. Oper. Res.* **16** 650–669.

——, ——. 1994. Finite master programs in regularized stochastic decomposition. *Math. Programming* **67** 143–168.

Jacobs, J., G. Freeman, J. Grygier, D. Morton, G. Schultz, K. Staschus, J. Stedinger. 1995. SOCRATES: a system for scheduling hydroelectric generation under uncertainty. *Ann. Oper. Res.* **59** 99–133.

Kall, P. 1988. Stochastic programming with recourse: upper bounds and moment problems—a review. J. Guddat, B. Bank, H. Hollatz, P. Kall, D. Klatte, B. Kummer, K. Lommatzch, K. Tammer, M. Vlach, K. Zimmermann, eds. *Advances in Mathematical Optimization.* Akademie-Verlag, Berlin, 86–103.

——, A. Ruszczyński, K. Frauendorfer. 1988. Approximation techniques in stochastic programming. Y. Ermoliev, R.J.-B. Wets, eds. *Numerical Techniques for Stochastic Optimization.* Springer-Verlag, Berlin, 33–64.

——, S.W. Wallace. 1994. *Stochastic Programming.* John Wiley and Sons, Chichester, UK.

King, A.J., R.J.-B. Wets. 1991. Epi-consistency of convex stochastic programs. *Stochastics* **34** 83–91.

Kusy, M.I., W.T. Ziemba. 1986. A bank asset and liability management model. *Oper. Res.* **34** 356–376.

Madansky, A. 1959. Bounds on the expectation of a convex function of a multivariate random variable. *Ann. Math. Statist.* **30** 743–746.

——. 1960. Inequalities for stochastic linear programming problems. *Management Sci.* **6** 197–204.

Mirchandani, P.B. 1976. Shortest distance and reliability of probabilistic networks. *Comput. Oper. Res.* **3** 347–355.

Mulvey, J.M., H. Vladimirou. 1989. Stochastic network optimization models for investment planning. *Ann. Oper. Res.* **20** 187–217.

OSL. 1993. *Optimization Subroutine Library.* International Business Machines, Kingston, NY.

Powell, W.B. 1986. A stochastic model of the dynamic vehicle allocation problem. *Transp. Sci.* **20** 117–129.

——, L.F. Frantzeskakis. 1994. Restricted recourse strategies for dynamic networks with random arc capacities. *Transp. Sci.* **28** 3–23.

Sen, S., R.D. Doverspike, S. Cosares. 1994. Network planning with random demand. *Telecommunication Syst.* **3** 11–30.

Stafford, M. 1997. A product-mix capacity planning model. Technical Report ORP97-03. Graduate Program in Operations Research, University of Texas, Austin, TX.

Vladimirou, H., S.A., Zenios, 1997. Stochastic linear programs with restricted recourse. *European J. Oper. Res.* **101** 177–192.

Wallace, S.W. 1987a. Investing in arcs in a network to maximize the expected max flow. *Networks* **17** 87–103.

——. 1987b. A piecewise linear upper bound on the network recourse function. *Math. Programming* **38** 133–146.

Wets, R.J.-B. 1996. Programming under uncertainty: the complete problem. *Zeitschrift für Warscheinlichkeitstheorie und Verw. Gebiete* **4** 316–339.

——. 1974. Stochastic programs with fixed recourse: the equivalent deterministic program. *SIAM Rev.* **16** 309–339.

——. 1983. Solving stochastic programs with simple recourse. *Stochastics* **10** 219–242.

Wollmer, R.D. 1985. Critical path planning under uncertainty. *Math. Programming Study* **25** 164–171.

——. 1991. Investments in stochastic maximum flow networks. *Ann. Oper. Res.* **31** 459–468.

Žáčková, J. 1966. On minimax solutions of stochastic linear programming problems. *Časopis pro Pěstování Matematiky* **91** 423–429.

Zangwill, W.I. 1967. Nonlinear programming via penalty functions. *Management Sci.* **13** 344–358.

Ziemba, W.T. 1974. Stochastic programs with simple recourse. P.L. Hammer, G. Zoutendijk, eds. *Mathematical Programming in Theory and Practice.* North-Holland, Amsterdam, 213–273.

Zipkin, P.H. 1980. Bounds for row-aggregation in linear programming. *Oper. Res.* **28** 903–916.